antiSMASH

Medema, Marnix H.; Blin, Kai; Cimermancic, Peter; de Jager, Victor; Zakrzewski, Piotr; Fischbach, Michael A.; Weber, Tilmann; Takano, Eriko; Breitling, Rainer

Published in:
Nucleic Acids Research

DOI:
10.1093/nar/gkr466

Publication date:
2011

Supplementary Tables I-III

Supplementary Table I: Signature HMMs for detection of secondary metabolite biosynthesis gene clusters

Compound class	Description	HMM name	Source
NRPS	Condensation domain	Condensation	PFAM PF00668.13
NRPS	Adenylation domain	AMP-binding	PFAM PF00501.21
NRPS	Adenylation domain with integrated oxidase	A-OX	This study
NRPS/PKS	Thiolation domain	PP-binding	PFAM PF00550.18
PKS	Ketosynthase domain	PKS_KS	SMART
PKS	Acyltransferase domain	PKS_AT	SMART
PKS	Trans-acyltransferase docking domain	ATd	This study
PKS (neg.)	Bacterial type I fatty acid synthase	bt1fas	This study
PKS (neg.)	Fungal type I fatty acid synthase	ft1fas	This study
PKS (neg.)	Type II fatty acid synthase	t2fas	This study
PKS (neg.)	FabH fatty acid synthase	fabH	This study
PKS	Enediyine ketosynthase	ene_KS	Yadav et al. (2009)
PKS	Modular ketosynthase	mod_KS	Yadav et al. (2009)
PKS	Hybrid ketosynthase	hyb_KS	Yadav et al. (2009)
PKS	Iterative ketosynthase	itr_KS	Yadav et al. (2009)
PKS	Trans-AT ketosynthase	tra_KS	Yadav et al. (2009)
PKS	Unusual PKS HglD-like	hglD	This study
PKS	Unusual PKS HglE-like	hglE	This study
PKS	Type II PKS ketosynthase	t2ks	This study
PKS	Type II PKS ketosynthase, model 2	t2ks2	This study
PKS	Type II PKS Chain length factor	t2clf	This study
PKS	Type III PKS N-terminal	Chal_sti_synt_N	PFAM PF00195.12
PKS	Type III PKS C-terminal	Chal_sti_synt_C	PFAM PF00195.12
Terpene	Terpene synthase	Terpene_synth_C	PFAM PF03936.9
Terpene	Phytoene_synthase	phytoene_synth	This study
Terpene	Lycopene cyclase	Lycopene_cycl	PFAM PF05834.5
Terpene	Terpene cyclase	terpene_cyclase	This study
Terpene	NapT7-like protein	NapT7	This study
Terpene	Fungal geranylgeranyl pyrophosphate synthase	fung_ggpp	This study
Terpene	Fungal geranylgeranyl pyrophosphate synthase, model 2	fung_ggpp2	This study
Terpene	Dimethylallyl tryptophan synthase	dmat	This study
Terpene	Trichodiene synthase	trichodiene_synt	This study

Lantibiotics	LanC-like lantibiotics biosynthesis protein	LANC_like	PFAM PF05147.6
Lantibiotics	Lantibiotic dehydratase, N-terminus	Lant_dehyd_N	PFAM PF04737.6
Lantibiotics	Lantibiotic dehydratase, C-terminus	Lant_dehyd_C	PFAM PF04738.6
Lantibiotics	Lantibiotic antimicrobial peptide 18	Antimicrobial18	PFAM PF08130.4
Lantibiotics	Gallidermin	Gallidermin	PFAM PF02052.8
Lantibiotics	Lantibiotic, type A	L_biotic_typeA	PFAM PF04604.6
Lantibiotics	Lantibiotic, gallidermin/nisin family	TIGR03731	TIGR03731
Lantibiotics	Lantibiotic leader lacticin 481 group	LE-LAC481	De Jong et al. (2010)
Lantibiotics	Lantibiotic leader mersacidin cinnamycin group	LE-MER+2PEP	De Jong et al. (2010)
Lantibiotics	Lantibiotic leader LanBC modified	LE-LanBC	De Jong et al. (2010)
Lantibiotics	Lantibiotic peptide lacticin 481 group	MA-LAC481	De Jong et al. (2010)
Lantibiotics	Lantibiotic peptide nisin epidermin group	MA-NIS+EPI	De Jong et al. (2010)
Lantibiotics	Lantibiotic peptide nisin group	MA-NIS	De Jong et al. (2010)
Lantibiotics	Lantibiotic peptide epidermin group	MA-EPI	De Jong et al. (2010)
Lantibiotics	Lantibiotic peptide two component alpha	MA-2PEPA	De Jong et al. (2010)
Lantibiotics	Lantibiotic peptide two component beta	MA-2PEPB	De Jong et al. (2010)
Lantibiotics	lantibiotic peptide lacticin 481 group (dufour et al)	LE-DUF	De Jong et al. (2010)
Lantibiotics	lantibiotic leader lacticin 481 group (dufour et al)	MA-DUF	De Jong et al. (2010)
Bacteriocin	Putative *Streptomyces* bacteriocin	strepbact	This study
Bacteriocin	Antimicrobial peptide 14	Antimicrobial14	PFAM PF08109.4
Bacteriocin	Bacteriocin_IIc domain	Bacteriocin_IIc	PFAM PF10439.2
Bacteriocin	Bacteriocin_IId domain	Bacteriocin_IId	PFAM PF09221.3
Bacteriocin	Bacteriocin_IIdc_cydomain	BacteriocIIc_cy	PFAM PF12173.1
Bacteriocin	Bacteriocin_II domain	Bacteriocin_II	PFAM PF01721.11
Bacteriocin	Bacteriocin_IIi domain	Bacteriocin_IIi	PFAM PF11758.1
Bacteriocin	Lactococcin	Lactococcin	PFAM PF04369.6
Bacteriocin	Antimicrobial peptide 17	Antimicrobial17	PFAM PF08129.4
Bacteriocin	Lactococcin 972	Lactococcin_972	PFAM PF09683.3
Bacteriocin	Lactococcin G-beta	LcnG-beta	PFAM PF11632.1
Bacteriocin	Subtilosin A	Subtilosin_A	PFAM PF11420.1
Bacteriocin	Cloacin	Cloacin	PFAM PF03515.7
Bacteriocin	Linocin M18	Linocin_M18	PFAM PF04454.5
Bacteriocin	Bacteriocin biosynthesis	TIGR03603	TIGR03603

	cyclodehydratase		
Bacteriocin	Bacteriocin biosynthesis docking scaffold	TIGR03604	TIGR03604
Bacteriocin	SagB-type dehydrogenase	TIGR03605	TIGR03605
Bacteriocin	Bacteriocin, circularin A/uberolysin famil	TIGR03651	TIGR03651
Bacteriocin	Bacteriocin, microcyclamide/patellamide family	TIGR03678	TIGR03678
Bacteriocin	Thiazole-containing bacteriocin maturation protei	TIGR03693	TIGR03693
Bacteriocin	Bacteriocin propeptide	TIGR03798	TIGR03798
Bacteriocin	Bacteriocin biosynthesis cyclodehydratase	TIGR03882	TIGR03882
Bacteriocin	Bacteriocin, BA_2677 family	TIGR03601	TIGR03601
Bacteriocin	Bacteriocin protoxin, streptolysin S family	TIGR03602	TIGR03602
Bacteriocin	Microviridin A	mvnA	This study
Bacteriocin	Thiostrepton-like thiopeptides	thiostrepton	This study
Beta-lactams	Beta-lactam synthase	BLS	This study
Beta-lactams	Clavulanic acid synthase-like	CAS	This study
Beta-lactams	Tabtoxin synthase-like	Tabtoxin	This study
Aminoglycosides / aminocyclitols	2-deoxy-scyllo-inososе synthase	DOIS	This study
Aminoglycosides / aminocyclitols	NeoL-like deacetylase	neoL_like	This study
Aminoglycosides / aminocyclitols	SpcD-/SpcK-like thymidylyltransferas	spcDK_like_glyc	This study
Aminoglycosides / aminocyclitols	SpcF-/SpcG-like glycosyltransferase	spcFG_like	This study
Aminoglycosides / aminocyclitols	StrH-like glycosyltransferase	strH_like	This study
Aminoglycosides / aminocyclitols	StrK-like phosphatase	strK_like	This study
Aminoglycosides / aminocyclitols	ValA-like 2-epi-5-epi-valiolone synthase	valA_like	This study
Aminoglycosides / aminocyclitols	2-epi-5-epi-valiolone synthase, SalQ-like	salQ	This study
Aminocoumarins	NovK-like reductase	novK	This study
Aminocoumarins	NovJ-like reductase	novJ	This study
Aminocoumarins	NovI-like cytochrome P450	novI	This study
Aminocoumarins	NovH-like protein	novH	This study
Aminocoumarins	SpcD/SpcK-like thymidylyltransferase, aminocoumarins group	spcDK_like_cou	This study
Siderophores	Siderophore synthase	IucA_IucC	PFAM PF04183.5
Ectoines	Ectoine synthase	ectoine_synt	This study

Butyrolactones	AfsA-like butyrolactone synthase	AfsA	PFAM PF03756.6
Indoles	StaD-like chromopyrrolic acid synthase domain	indsynth	This study
Nucleosides	LipM-like nucleotidyltransferase	LipM	This study
Nucleosides	LipU-like protein	LipU	This study
Nucleosides	LipV-like dehydrogenase	LipV	This study
Nucleosides	ToyB-like synthase	ToyB	This study
Nucleosides	TunD-like putative N-acetylglucosamine transferase	TunD	This study
Nucleosides	Pur6-like synthetase	pur6	This study
Nucleosides	Pur10-like oxidoreductase	pur10	This study
Nucleosides	NikJ-like protein	nikJ	This study
Nucleosides	NikO-like enolpyruvyl transferase	nikO	This study
Phosphoglycolipids	MoeO5-like prenyl-3-phosphoglycerate synthase	MoeO5	This study
Melanins	MelC-like melanin synthase	melC	This study
Others	NAD-binding domain 4	NAD_binding_4	PFAM PF07993.5
Others	LmbU-like protein	LmbU	This study
Others	Goadsporin-like protein	goadsporin_like	This study
Others	Neocarzinostatin-like protein	Neocarzinostat	This study
Others	Cyanobactin protease	cyanobactin_synth	This study
Others	Cyclodipeptide synthase	cycdipepsynth	This study
Others	Fom1-like phosphomutase	fom1	This study
Others	BcpB-like phosphomutase	bcpB	This study
Others	FrbD-like phosphomutase	frbD	This study
Others	MitE-like CoA-ligase	mitE	This study
Others	Valanimycin biosynthesis VlmB domain	vlmB	This study
Others	Pyrrolnitrin biosynthesis PrnB domain	prnB	This study

Supplementary Table II: Rules for detection of secondary metabolite biosynthesis gene clusters

Biosynthetic class	Rules
Type I PKS	- KS & AT HMM scores > 50 within one protein - KS score > bactTypeIFAS / fungTypeIFAS / HglD&E / FabH scores
Trans-AT type I PKS	- trans-AT docking domain HMM score > 65 - KS score > 50 - no match to rules for normal type I PKSs as above
Type II PKS	- type II KS or CLF HMM score > 50 - KS/CLF score > enedyineKS / modularKS / hybridKS / iterativeKS / transATKS / bactTypeIFAS / fungTypeIFAS / TypeIIFAS / HglD&E / FabH HMM scores - no match to rules for normal/trans type I PKSs as above
Type III PKS	- Chal_sti_synt_C or Chal_sti_synt_N HMM scores > 35 - no match to rules for normal type I&II PKSs as above
Unusual HglD/E-like PKS	- HglE or HglD HMM scores > 50 - HglD/E HMM score > bactTypeIFAS / fungTypeIFAS / TypeIIFAS / FabH HMM scores - no match to rules for normal type I&II&III PKSs as above
Non-ribosomal peptide synthetase	- C & A / A-OX domain HMM scores > 20 within one protein *OR* single domain C & A proteins scores > 20 within 20 kb distance
Terpene	- Terpene_Synth HMM score > 23 *OR* phytoene_synt HMM score > 20 *OR* Lycopene_cycl HMM score > 80 *OR* terpene_cyclase HMM score > 50 *OR* NapT7 HMM score > 250 *OR* fung_ggpps HMM score > 420 *OR* fung_ggpps2 HMM score > 312 *OR* dmat HMM score > 200 *OR* trichodiene_synth HMM score > 150
Lantibiotics	- LANC_like HMM score > 80 *OR* (Lant_dehydN and LantdehydC HMM scores > 20 within one protein) *OR* one of a range of lantibiotic prepeptide HMM scores > 20 *OR* TIGR03731 HMM score > 18
Bacteriocins	- Strepbact HMM score > 50 *OR* Antimicrobial14 HMM score > 90 *OR* Bacteriocin_IId HMM score > 23 *OR* BacteriocIIc_cy HMM score > 92 *OR* Bacteriocin_II HMM score > 40 *OR* Lactococcin HMM score > 24 *OR* Antimicrobial17 HMM score > 31 *OR* Lactococcin_972 HMM score > 25 *OR* Bacteriocin_IIc HMM score > 27 *OR* LcnG-beta HMM score > 78 *OR* Bacteriocin_Iii HMM score > 56 *OR* Subtilosin_A HMM score > 98 *OR* Cloacin HMM score > 27 *OR* Linocin_M18 HMM score > 25 *OR* TIGR03603 HMM score > 150 *OR* TIGR03604 HMM score > 440 *OR* TIGR03605 HMM score > 200 *OR* TIGR03651 HMM score > 18 *OR* TIGR03678 HMM score > 35 *OR* TIGR03693 HMM score > 400 *OR* TIGR03798 HMM score > 16 *OR* TIGR03882 HMM score > 150 *OR* TIGR03601 HMM score > 50 *OR* TIGR03602 HMM

		score > 50 *OR* mvnA HMM score > 20 *OR* thiostrepton HMM score > 20
Beta-lactams	-	Beta-lactam synthase HMM score > 250 *OR* clavulanic acid synthase HMM score > 250 *OR* tabtoxin synthase score > 500
Aminoglycosides / aminocyclitols	-	strH HMM score > 50 *OR* strK1 HMM score > 800 *OR* strK2 HMM score > 650 *OR* NeoL HMM score > 50 *OR* DOIS HMM score > 500 *OR* ValA HMM score > 600 *OR* SpcFG HMM score > 200 *OR* SpcDK_glyc HMM score > 600 *OR* salQ HMM score > 480
Aminocoumarins	-	novK HMM score > 200 *OR* novJ HMM score > 350 *OR* novI HMM score > 600 *OR* novH HMM score > 750 *OR* spcDK_like_cou HMM score > 600
Siderophores	-	IucA_IucC HMM score > 30
Ectoines	-	Ectoine_synt HMM score > 35
Butyrolactones	-	AfsA HMM score > 25
Indoles	-	ind_synth HMM score > 100
Nucleosides	-	LipM HMM HMM score > 50 *OR* LipU HMM HMM score > 30 *OR* LipV HMM HMM score > 375 *OR* LipU HMM HMM score > 30 *OR* ToyB HMM HMM score > 175 *OR* TunD HMM HMM score > 200 *OR* pur6 HMM HMM score > 200 *OR* pur10 HMM HMM score > 600 *OR* nikJ HMM HMM score > 200 *OR* nikO HMM HMM score > 400
Phosphoglycolipids	-	MoeO5 HMM score > 65
Melanins	-	melC HMM score > 40
Others	-	PP-binding & AMP-binding HMM scores > 20 within one protein *OR* (PP-binding HMM score > 20 and A-OX HMM score > 50 within one protein) *OR* (NAD_binding_4 HMM score > 40 and A-OX HMM score > 50 within one protein) *OR* (NAD_binding_4 HMM score > 40 and AMP-binding HMM score > 20 within one protein) *OR* LmbU HMM score > 50 *OR* goadsporin_like HMM score > 500 *OR* Neocarzinostat HMM score > 28 *OR* cyanobactin_synth HMM score > 80 *OR* cycdipepsynth HMM score > 110 *OR* fom1 HMM score > 750 *OR* bcpB HMM score > 400 *OR* frbD HMM score > 350 *OR* mitE HMM score > 400 *OR* vlmB HMM score > 250 *OR* prnB HMM score > 200

Supplementary Table III: HMM library for PKS/NRPS domain architecture analysis

Compound class	Description	HMM name	Source
NRPS	Condensation domain	Condensation	PFAM PF00668.13
NRPS	Condensation domain that links L-amino acid to peptide ending with D-amino acid	Condensation_DCL	Rausch et al. (2007)
NRPS	Condensation domain that links L-amino acid to peptide ending with L-amino acid	Condensation_LCL	Rausch et al. (2007)
NRPS	Dual condensation / epimerization domain	Condensation_Dual	Rausch et al. (2007)
NRPS	Starter condensation domain	Condensation_Starter	Rausch et al. (2007)
NRPS	Putatively inactive glycopeptide condensation-like domain	CXglyc	This study
NRPS	Glycopeptide condensation domain	Cglyc	This study
NRPS	Heterocyclization domain	Heterocyclization	This study
NRPS	Epimerization domain	Epimerization	Weber et al. (2009)
NRPS	Adenylation domain	AMP-binding	PFAM PF00501.21
NRPS	Adenylation domain with integrated oxidase	A-OX	This study
NRPS	Peptidyl-carrier protein domain	PCP	This study
NRPS	NRPS COM domain N-terminal	NRPS-COM_Nterm	This study
NRPS	NRPS COM domain C-terminal	NRPS-COM_Cterm	This study
NRPS	4'-phosphopantetheinyl transferase	ACPS	PFAM PF01648.9
PKS	Ketosynthase domain	PKS_KS	SMART
PKS	Acyltransferase domain	PKS_AT	SMART
PKS	Ketoreductase domain	PKS_KR	SMART
PKS	Enoylreductase domain	PKS_ER	SMART
PKS	Dehydratase domain	PKS_DH	SMART
PKS	Dehydratase domain of trans-AT PKSs	PKS_DHt	This study
PKS	Acyl-carrier protein domain	ACP	This study
PKS	Trans-acyltransferase docking domain	Atd	This study
PKS	Enediyine ketosynthase	Ene_KS	Yadav et al. (2009)
PKS	Modular ketosynthase	Mod_KS	Yadav et al. (2009)
PKS	Hybrid ketosynthase	Hyb_KS	Yadav et al. (2009)
PKS	Iterative ketosynthase	Itr_KS	Yadav et al. (2009)
PKS	Trans-AT ketosynthase	Tra_KS	Yadav et al. (2009)
PKS	Polyketide cyclase	Polyketide_cyc	PFAM PF03364.13
PKS	Polyketide cyclase /	Polyketide_cyc2	PFAM PF10604.2

	dehydratase		
PKS	PKS N-terminal docking domain	PKS_Docking_Nterm	This study
PKS	PKS C-terminal docking domain	PKS_Docking_Cterm	This study
PKS	Co-enzyme A ligase domain	CAL	This study
NRPS/PKS	C-methyl transferase	cMT	Ansari et al. (2008)
NRPS/PKS	O-methyl transferase	oMT	Ansari et al. (2008)
NRPS/PKS	N-methyl transferase	nMT	Ansari et al. (2008)
NRPS/PKS	Aminotransferase class I&II	Aminotran_1_2	PFAM PF00155.14
NRPS/PKS	Aminotransferase class III	Aminotran_3	PFAM PF00202.14
NRPS/PKS	Aminotransferase class IV	Aminotran_4	PFAM PF01063.12
NRPS/PKS	Aminotransferase class V	Aminotran_5	PFAM PF00266.12
NRPS/PKS	Thiolation domain	PP-binding	PFAM PF00550.18
NRPS/PKS	Thioesterase domain	Thioesterase	PFAM PF00975.13
NRPS/PKS	Terminal reductase domain	TD	This study